Journal of Theoretical and Applied Information Technology
15th June 2022. Vol.100. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HOW MARKETERS CAN INCREASE THE RELEVANCE OF EMAIL MARKETING CAMPAIGNS: DATA ANALYSIS WITH MACHINE LEARNING METHODS

[1]REDOUAN ABAKOUY, [2]EL MOKHTAR EN-NAIMI, [3]ANASS EL HADDADI, [4]LOTFI ELAACHAK

[1,2,4] LIST Laboratory, Department of Computer Science, the Faculty of Sciences and Technologies, UAE, Tangier, Morocco,
[3] DMI Laboratory, ENSAH, Abdelmalek Essaadi University, Al Hoceima, Morocco
E-mail: [1] redouan.abakouy@gmail.com, [2]ennaimi@gmail.com, [3]anass.elhaddadi@gmail.com, [4]lotfi1002@gmail.com

ABSTRACT

Despite the emergence of many digital marketing strategies, email remains the most effective strategy of all. It generates a high return on investment (ROI) and offers a fast way of retention of customers. Is still the preferred channel, not only for businesses, but also for clients, Convinced of its great capacity for influence, more and more companies are looking for a relevant marketing strategy.
Convincing the customer to open the email is the first step to an effective campaign. Therefore, it is important to understand how marketers can improve the open rate of a marketing campaign. Data Driven Marketing encompasses the many techniques and applications that affect digital marketing directly related to the use of customer data. Machine learning can predict user needs based on customer data and past behaviors. Those predictions can then be used to suggest offers that are based on the individual. The objective of the work is to analyze the main factors that increase the opening rate of marketing campaigns. To this end, we are developing classification algorithms to predict whether a campaign will be classified as relevant or irrelevant. To achieve this, we used and evaluated three different classifiers. Our results showed that it is possible to predict the performance of a campaign with an accuracy of about 82%, using Adaptive Boosting algorithm and the redundant filter selection technique.

Keywords: *Digital Marketing; Email Marketing; Machine learning, Data Analysis*

1. INTRODUCTION

Email Marketing is one of the most important channel in Digital Marketing. It yields a high return on investment (ROI) for the company and offers a cheap and fast way to reach existent or potential clients. Motivate clients to open the email is the first step for a relevant campaign. Thus, it is important to understand how marketers can improve the open rate of a marketing campaign [1].

Marketing is a key part of business operations. Therefore, a success or a failure of a modern company is considerably based on a quality of its marketing campaigns.

According to a Forrester study released in late 2013, consumers were more than twice as likely to delete most emails without reading them (between 59% and 73% of emails are deleted without being read)[16]. Launching a successful marketing campaign in a very competitive market segment, is evidently a challenging job [5].

To strengthen the connection with the customer, it is important to create a feeling of personal communication via email. Simply addressing them by name in subject lines makes your customers feel appreciated and increases ROI by 26%. The content must be relevant. For example, it will be very useful to know the gender and the location of the customers, as well as the history of clicks and purchases of the

Journal of Theoretical and Applied Information Technology
15th June 2022. Vol.100. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

customers.

To craft relevant offers, marketers must understand customer preferences and interests,

The “gut-feeling” [3] is not an effective decision-making process, so, Market-ing is a complex business, and results and decisions can’t be measured with the gut. Sure, there must be people out there who can do either accurately, but with-out any means of measuring a decision-making process, success could be just a coincidence. Business analytics revolutionizes the face of decision-making sup-port. Today, we observe an exceptional attention in quantitative decision aids and analytic models. Campaign planners require to answer three questions: when to send an offer (timing), how often to send an offer (frequency) and target group selection [6].

Business analytics revolutionizes the face of decision-making support. Today, we observe an exceptional attention in quantitative decision aids and analytic models. Campaign planners require to answer three questions: when to send an offer (timing), how often to send an offer (frequency) and target group selection [9]. Such approach allows to customize the product properties per customer basis, thus increasing campaign efficiency and reducing costs[10].

By using data-driven approach[3,12,13,14,22], marketers can predict the performance of a campaign before even sending it. In fact, if a marketer knows in advance if a campaign is going to be successful or not, it provides the opportunity to sooner correct problems that could strongly impact its revenue. To our knowledge, this is the first publication that does an extensive qualitative analysis of the main factors driving the opening behavior of financial marketing campaigns. Nowadays, financial institutions are using Email Marketing as an important source to reach their clients. The major problem with email marketing is low customer engagement with received messages, thus majority of emails are deleted without being read[16]. Understanding the data and numbers at the heart of marketing strategy and decision-making can enable marketers to predict marketing campaigns based on customer shopping experience and personalized engagement. Thus, Some of the biggest challenges for email marketing are provoking engagement, increasing conversions, and generating leads. In order to improve these metrics, and have a successful email marketing strategy, Marketers need to collect and analyze data. These include subject lines, frequency, day and time emails are sent, and the type of content and to successfully develop email campaigns, they must ensure the ongoing collection and effective use of customer data.

that drives client engagement. this work will guide marketers on how to implement successful campaigns in this field.

During this research work, we study what are the main factors driving the open rate of email marketing campaigns. Motivate the recipients to open the email is the first step for an effective campaign, since it determines the reach of the cam-paign. Therefore, it’s important for marketers to first understand how they can improve the email open rate. With that purpose, our research team has deployed three machine learning models for prediction including: Decision tree[2], Bagging classifier[15] and Adaptive Boosting[11] with the aim to fully evaluate the prediction models already established, the three models have been trained according to different combinations of scenarios. we developed a classification algorithm that can accurately predict if a campaign will be classified as relevant or irrelevant. A campaign is labeled as relevant if it has an open rate higher than the aver-age, otherwise it is classified as irrelevant. Additionally, we did a text analysis of the subject line and preheader to discover which keywords and keyword combinations are associated with a higher email open rate. To validate the results obtained in a real setting, we performed A/B testing in the deployment stage.

The remaining part of this paper is arranged as follows: Section 2 recaps the related works in the area of marketing; Section 3 presents business understanding; Section 4 presents data-driven approach, the features used, how they are is collected, processed and analyzed. Section 5 presents data exploration, how to fix what are the most appropriate features to the target variable, the feature selection experiment performed was to filter the input unnecessary features; Section 6 shows the three classification models that are employed for the prediction purposes in this research. results are discussed in Section 7. Finally, conclusions are presented and the future work is resumed in Section 8.

2. RELATED WORK

Machine-learning techniques and classification algorithms have the potential to improve business processes, predict future outcomes, and save money. While there are many different machine-learning algorithms, they can be split into two different categories:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

supervised and unsupervised. These categories differ based on what criteria they use to classify data or predict outcomes.

The current research studies in email marketing are based on the features extracted from the campaigns and email recipients profiles. In this work, because our objective was to analyze qualitatively the main factors contributing to mar-keting campaigns with an email open rate above the average.

In 2014, Raju Balakrishnan and Rajesh Parekh, have proposed a method to help the editors by predicting subject line open rates by learning from past subject lines. They used syntactical, historical and derived features of each key-word in the subject line and of the entire subject line. A random forest is trained to integrate these features to predict the performance. they used a dataset of more than a hundred thousand different subject lines with many billions of im-pressions to train and test the method. For the baseline, the open rate prediction was equal to the mean open rate of past emails that used the same subject line. For new subject lines, the open rate was predicted as the average open rate of all the subject lines[4].

In 2014, Using metadata and social network analysis, Pål Sundsøy, Johannes Bjelland, Asif M Iqbal, Alex Sandy Pentland, Yves-Alexandre de Montjoye, cre-ated new metrics to identify customers that are the most likely to convert into mobile internet users. These metrics fall into 3 categories: discretionary income, timing, and social learning. Using historical data, a machine learning prediction model is then trained, validated, and used to select a treatment group. Experi-mental results with 250 000 customers show a 13 times better conversion-rate than control group. The control group is selected using the best practice market-ing. The model also shows very good properties in the longer term, as 98% of the converted customers in the treatment group renew their mobile internet packages after the campaign, compared to 37% in the control group. These results show that data-driven marketing can remarkably ameliorate conversion rates over cur-rent best-practice marketing strategies. This study, based on several modeling algorithms such as support vector machine, neural networks and decision tree to classify natural converters. The bagging decision tree however turned out to be more stable when tested across different samples[3].

In 2015, Luo et al. developed a classification algorithm to predict if a targeted email will be open or not. For each email recipient, the model classified the email in “open” or “unopen”. The model used features extracted from the emails and from the email recipients’ profiles. For the prediction phase they used two differ-ent classifiers, Support Vector Machine and Decision Tree, on two different da-tasets using different feature selection methods (include or not include the recip-ient´s domains). The Decision Tree outputted the other classifier, achieving a F1-measure rate of approximately 80% on the “opens”, in the case of considering all features. In the case where the recipient´s domains were not considered, the per-formance of both classifiers dropped, which indicates this component is im-portant to predict the email open behavior[2].

In 2015, Shaokun Fan, Raymond Y.K. Lau and J. Leon Zhao have proposed to use a marketing mix framework for guiding research in big data management for marketing intelligence. they have identified the data sources, methods, and appli-cations in different marketing perspectives. They have discussed the challenging issues related to big data management in the context of diverse marketing per-spectives. They also have underlined the future areas of research for large data management [8].

In 2018, Jaidka et al. also studied the problem of predicting email opens, based on the subject line. They explored the differences in the recipient´s preferences for subject lines sent by different business areas (Finance, Cosmetics and Television). The methodology used was a Data Mining model to predict the open rate of different email subject lines, a regression analysis to study the effect of different subject line language styles in the open rate and a domain adaptation method. The learning model used was a five-fold cross–validated weighted linear regression, which predictions improved over the baselines - state-of-the-art model and the mean open rate of the entire dataset. The use of the domain adaptation meth-od improved the prediction of the model for unseen domains and business. They concluded that using certain styling strategies in the subject line, according to the business area of the campaign, can strongly impact the email open rate[7].

In 2019, Andreia Conceição and João Gama, analyzed the factors driving the open rate of financial email marketing campaigns. they developed a classification algorithm that can accurately predict if a campaign will be labeled as Successful or Failure. The results indicated that it is possible to predict the success of a

Journal of Theoretical and Applied Information Technology
15th June 2022. Vol.100. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

campaign with roughly 82% accuracy, by using the Random Forest algorithm and the redundant filter selection technique. In addition, a text analysis of the subject line and preheader was made to find which keywords and keyword combinations provoke a best open rate. The results were then validated in a real setting through A/B testing[1].

The contributions of our work to these papers are the inclusion of the preheader and the email sender as features in the prediction task. Before opening an email, the recipient has also information of these components; therefore, we decided to test if these features are important to predict the open behavior of a campaign.

3. BUSINESS UNDERSTANDING

The goal of company is to understand how they could improve the open, click and conversion rates of their email marketing campaigns. To meet this business goal, we have developed classifica-tion algorithms that can accurately predict whether a campaign will be classified as relevant or irrelevant. Additionally, we applied a subject line and preheader text analysis to find which keywords and keyword combinations trigger a higher open rate. To validate the results obtained in the company's activity, we carried out A / B tests.

4. DATA-DRIVEN APPROACH

Marketing science[19] has a long practice of implementation new challenges, new methods, and new disciplines. In this research work, we collected massive real data to study and analyze the use of email marketing for this purpose. Our datasets include history of customer actions such as opening an email or clicking on an email or email lead. Based on the collected data, we extract campaign characteristics and build customer profiles.

The features used in this work include the subject-lines, from-lines, offers, preheader, vertical and time when an email is opened. Additionally, the recipient profile features employed in this work include IP, the location of client, operating system,
device, the response time and the email domain. Though, it is worth mentioning that some of the features can only be collected when there is an action on the email.

4.1 Extraction of campaign Features

In this study, we have collected many features from each campaign. These are: the words of the subject-line, number of keywords; subject length, the words of the from-line, preheder, Email content, offers, vertical (as mentioned in Fig. 2), geolocation[17-18], the day the email is sent and the time the email is opened, number of emails sent; number of emails delivered; number of emails opened, The time and the day are two features that are very straight forward. For the time feature, we divided the time of a day into four parts: T1 (00:00–06:00), T2 (06:00–12:00), T3 (12:00–18:00), T4 (18:00–24:00).

Prediction based on the individual words within the subject line and preheader of an email could be very demanding, because the number of words in a subject line and preheader is not that many to start with. In the latter case, it is difficult to associate that word with the open rate of an email. Generally, such an email contains number of components such as, Email sender, Subject line and Email content (Generally contains the preheader and number of deal links with pictures and descriptions of deals), as shown in Figure 1:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 1: A Mail View In The Inbox And After Opening The Mail. Email Sender, Subject Line And Email Content Are The Prominent Components The Client Would See In His Inbox.

The variables used for this study are described in Table 1. These variables were extracted and derived from email campaigns of the studied company.

Table 1: Features For Campaigns

Features	Description
Delivered email	Number of emails sent
Bounced emails	Number of emails bounced
Deferred emails	Number of emails deferred
Sent days	Number of days the campaign was sent
Open days	The days when the email is opened by the customer
Opens timing	The time the email is opened : T1 (00:00–06:00), T2 (06:00–12:00), T3 (12:00–18:00), T4 (18:00– 24:00)
Vertical	Beauty, cash, health, legal, vacance, job ...
Number of keywords	The total number of keywords in the email subject line and
Length subject line	The number of characters in the subject line
Length preheader	The number of characters in the preheader
Personalization	If the email subject line and/or preheader has a personalized greeting
Has Digits	Whether subject line contains digits (0-9).
Has punctuation	Whether subject line contains punctuation marks
Open Rate	Open Rate
Occurrence score first, second, third and fourth keyword	Occurrence score of the first, second, third and fourth keyword of the subject line and preheader, counting from the left to the right
Classification	Categories: • Relevant – if the open rate is above the average; • Irrelevant– if the open rate is below the average; The email open rate of a campaign is calculated by dividing the number of emails opened by the number of emails delivered.

Table 2. Categories Of The Words Used In The Subject Line Of An Email[2]

Category	Description	Examples
Client	Related to the business name of the	McDonalds
Business	Related to the business product of the	Hamburg, Fries, Pop, . . .
Time	A name of a time or day	Holidays, July, Weekend, .
Location	A name of a location	New York, Amherst, . . .
Highlight	An adjective to highlight information	Happy, Brilliant, . . .
CAP	Word in the upper case form	COUPON, THURSDAY, .

Figure 2: Opening Emails By Vertical

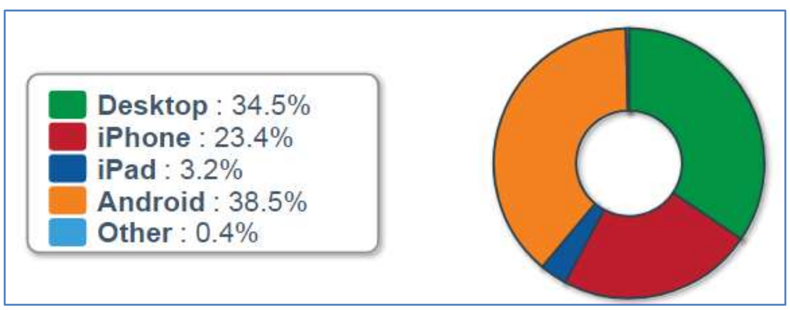

Figure 3: Percentage Of Open Instances On Different Device Types

It's usually difficult to predict email open rates based on subject lines. Depending on the number of times a subject line has been used in the past, estimating its performance can be from easy to very difficult. On the easy side, if an exact subject line has been used in the past in a large number of email campaigns, the open rates of the subject line gives us a statistically significant estimate of how well the subject line will be performing. On the other hand, if the subject line has not been used in the past, it is very difficult to predict open rates. Between these two extreme cases, subject lines that are rarely used in the past and those similar to the previously used subject have varying degrees of difficulty predicting open rates.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

We indicate the subject lines as an organized composition of keywords. The features are derived at a keyword level, and the order of the keywords is taken into consideration for prediction. In this study, the problem may be fixed as to predict the open-rate OPi of emails with the subject line SUBi where, SUBi = (KW1, KW2, .., KWn) and KWj is the jth keyword in SUBi and n is the total number of keywords. To predict the open rates, we extract the features keywords in the subject line, and the subject line as a full.

The occurrence score of a keyword is equal to the difference between the number of times the keyword was in the "Relevant" set in past campaigns and the number of times it was in the "Irrelevant" set [2]. This feature captures the consistency in the performance of a keyword. In fact, a positive occurrence score indicates the keyword has a good performance, as it was principally existing in campaigns with an open rate above the average.

4.2 Client profile features collection

In addition to the characteristics of the campaign, the properties of email client are also very relevant in prediction if the email will be opened and clicked by the recipient or not. Researches show that the geographical location is important for many internet applications. It helps to understand the customer distribution and enables location-based advertising services. Thus, in this work, we explored whether the same will apply in the email marketing field too. Many researches have been done on identifying the geolocation from IP addresses. We employed the ip2location database https://www.ip2location.com, to identify the country, the state and the city of the client based on his/her IP address, and include this as one of the features of the recipient. We collect the OS type, the Device type, the Browser type and Domain of the recipient that plays an important role in predicting the open and click rates. Table 3 presents the features used for customer profiling.

Table 3: Features For Email Customer Profiling

Type	Data source	Examples
Country	IP address	US, UK, NL, AU, . . .
State	IP address	California, . . .
OS type	User-Agent string	iOS, Android, . . .
Device type	User-Agent string	iPad, PC, . . .
Browser type	User-Agent string	Chrome, ...
Domain	The domain part of recipient's email address	yahoo.com, gmail.com, aol.com, hotmail.com, outlook.com, Verizon.net, icloud.com . . .

5. DATA EXPLORATION

This work aims to invent how marketers can improve the email open rate of their marketing campaigns. Thus, in this study, the aim was to find the variables significantly associated with the open rate. Text analytics is the practice of extracting meaning out of text. The goal of this text analysis was to discover which keywords and keyword combinations are associated with a higher open rate and click rate. To derive the set of keyword combinations we have extracted the first right of each keyword in the email subject line and preheader. The Feature Selection process has a huge impact on the performance of an algorithm.

Hence, it is important to fix what are the most appropriate features to the target variable. To this end, the feature selection experiment performed was to filter the input unnecessary features. The presence of redundant features does not add any meaningful information to the existing feature set. Therefore, we can remove one of two highly correlated variables without losing important data. By reducing the set of features, the running time of the algorithm considerably decreases, and, at the same time, the performance of the model increases [1].

6. MODELLING

In this research, Our data sets include a collection history of customer actions such as opening an email or clicking on an email, or making purchases. This process has allowed us

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to extract the features of the campaigns and build the profiles of the clients.

The email marketing features employed in this work include the subject-lines, from-lines, offers, vertical and time when an email is opened. In addition, the recipient profile features employed in this work include the location of the client, the OS type, the Device type, the Browser type, Domain of the recipient and the response time. Though, it is worth mentioning that some of the features can only be collected when there is an action on the email.

In this research work, classification algorithms were applied because the Data Mining problem in hand was to find to which set of classes (relevant or irrelevant) a new campaign belongs to, based on the training set containing past campaigns whose classification is already known. The classification algorithms used were the following: Decision Tree, Bagging classifier and Adaptive Boosting[11]. Each algorithm has its pros and cons, but the most important thing is that all of them require a well featured dataset for the creation of the model that will be able to classify

accurately the data test. To validate the performance of these models in new and unseen data, the 10-Fold Cross-Validation method was used.

The comparison of the results presented in table 4, permitted us to notice that the model which had the best performance was the Adaptive Boosting, when using the redundant feature selection technique. This model accurately predicted 82% of the observations, achieving approximately an AUC of 71% and F1-score of 74% (for the relevant class). The recall for the relevant class was slightly lower.

Table 4. Machine Learning Algorithms For Models Establishment

Algorithm	Accuracy	AUC	F-S	Amount of data
Decision tree	0.74	0.61	0.74	1000000
Bagging classifier	0.75	0.66	0.74	1000000
Adaptive Boosting	0.82	071	0.74	1000000

7. RESULTS AND DISCUSSION

During this study, three machine learning algorithms have been tested s to classify clickers. Often it is difficult to define if one algorithm is better than the other, because the judgments depends on several factors including accuracy, type of the data treated, training time… The final model is a combination between accuracy and stability where, based on earlier experience, we put more weight on stability. The Adaptive Boosting classifier however turned out to be more stable when tested across different samples.

In our case based on F1-score which represents the weighting of both factors (recall & precision) even on the rock curve. To evaluate each used algorithm we have used AUC-ROC curve, which is a measure of performance for a classification problem at different threshold settings. OCR is a probability curve and the AUC represents the degree or extent of separability. It shows how many models are able to differentiate classes. The higher the value of the AUC, the more the model is able to distinguish between classes (a good prediction). The ROC curve is plotted with TPR (true positive rate) versus FPR (positive false rate), where TPR is on the y-axis and FPR is on the x-axis.

According to the four coefficients “True Positive, False Positive, True Negative, False Negative,”of the confusion matrix it’s possible to calculate precision, recall, F1-score, accuracy, etc:

Accuracy (1)- is the most intuitive performance measure. It is simply a ratio of correctly predicted observations to the total number of observations. the accuracy is greater than the model is better.

Accuracy = (TP + TN) / (TP + FP + FN + TN) (1)

Precision (2)- Precision is the ratio of the correctly predicted positive observations to the total predicted positive observations. Precision answers the next question On all positive classes, how much have we predicted correctly. It should be as high as possible.

Precision = TP / (TP + FP) (2)

Recall (3)- The reminder is the relationship between the correctly predicted positive observations and all observations of the real class.

Recall = TP / TP + FN (3)

F1 Score (4)- The F1 score is the weighted average of accuracy and recall. Therefore, this score takes into account both false positives and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

false negatives.

F1-score = 2 * (Recall * Precision) / (Recall + Precision) (4)

In our case, the OCR curve of the Decision Tree algorithm, The AUC is equal to 0.61, which means that there is a 61% chance that the model can distinguish between positive and negative classes.

Finally, and After the fit and training of the three algorithms cited above, and according to the obtained results shown bellow in tables 4. we can see the Adaptive Boosting classifier performs better in all the scenarios. However, the parameter sensitivity of these two algorithms for this task is left for future work.

We examined all the features, and identified that a high percentage of the features are clientid, vertical, from-lines, subject-lines, device, offer, timing.... The results based upon the three traditional evaluation measurements, namely: Accuracy, AUC and F1-Score are shown in table 4.

During this work, we encountered difficulties related to the deliverability of marketing campaigns, so some of messages do not arrive in the client's inbox, this influenced the result obtained, for this purpose, improving deliverability can allow us to obtain a better correlation of data thus better results.

8. CONCLUSION

Email marketing is a powerful marketing channel, it can be more effective if emails can reach the right customers at the right time. Personalization and relevance have become more and more critical factors for all kinds of media used in digital marketing[21]. Through this work, we employed a learning model for predicting the "relevant" and "irrelevant" email marketing campaigns. The model is based on the features collected from an opened email and campaigns. The features that revealed to be important to predict the opening performance of a campaign were the number of emails sent, the length of the preheader and the occurrence score of the keywords used in the subject line and preheader. Regarding the email subject line, we advise marketers to avoid using long subject lines since it can negatively impact the open rate.

With that purpose, we developed a classification algorithms that can predict if a campaign will be labeled as Relevant or Irrelevant. A campaign is classified as Relevant if it has an open rate above the average, else it is labeled as Irrelevant. We tested three different classifiers Decision tree, Bagging classifier and Adaptive Boosting. The model that achieved the best performance was the Adaptive Boosting, when using the redundant filter selection technique. This model could accurately predict the performance of 82% campaigns, achieving an AUC of 71% and a F-score of 74% (for the Relevant class). By using this model, marketers will have the chance to sooner correct potential problems in a campaign that could highly impact its revenue.

During the phases of this research, we encountered difficulties related to the deliverability of the marketing campaigns, so some of the messages do not arrive in the client's inbox, this problem influenced the result obtained, for this purpose, we will working with marketers for improving deliverability which can allow us to get better correlation of data and better results.

In the future, we challenge to integrate marketing automation, this can save a significant amount of time and money. Therefore, a generated email campaign can be set up much faster with a smaller margin of error and fewer team members. Thus more sophisticated profiles can be created to include email features when clients have "clicks" on the email or purchase, thus more personalization and relevance, and the profile can be ever adaptive and changing to reflect customer behavior and preferences. Finally, in order to increase the performance of analysis and decision-making in real-time, for the benefit of marketers, we will integrate the BIG DATA ecosystem.

REFERENCES:

[1] A. Conceição, J. Gama: "Main Factors Driving the Open Rate of Email Marketing Campaigns". P. Kralj Novak et al. (Eds.): DS 2019, LNAI 11828, pp. 145–154, 2019. Springer Nature Switzerland AG 2019. https://doi.org/10.1007/978-3-030-33778-0_12

[2] Luo, X., Nadanasabapathy, R., Zincir-Heywood, A.Nur, Gallant, K., Peduruge, J.: "Predictive analysis on tracking emails for targeted marketing". In: Japkowicz, N., Matwin, S. (eds.) DS 2015. LNCS (LNAI), vol. 9356, pp. 116–130. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24282-8_11

[3] P. Sundsøy, J. Bjelland, A. M. Iqbal, A. S. Pentland and Y-A. d. Montjoye. 2014. "Big Data-Driven Marketing: How machine learning outperforms marketers' gut-feeling", Social

Computing, Behavioral-Cultural Modeling & Prediction. Lecture Notes in Computer Science, 8393 (2014), 367-374Ding, W. and Marchionini, G. 1997. A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[4] Balakrishnan, R., Parekh, R.: "Learning to predict subject-line opens for large-scale email marketing". In: 2014 IEEE International Conference on Big Data (Big Data), pp. 579–584. IEEE, October 2014

[5] P. Łady ˙zy ´nski, K. ˙Zbikowski, P. Gawrysiak, 2019, "Direct marketing campaigns in retail banking with the use of deep learning and random forests", Expert Systems With Applications 134 (2019) 28–35

[6] S. Lessmann, J. Haupt, K. Coussement, K. W. De Bock, 2019, "Targeting customers for profit: An ensemble learning framework to support marketing decision-making", https://doi.org/10.1016/j.ins.2019.05.027

[7] Jaidka, K., Goyal, T., Chhaya, N.: "Predicting email and article clickthroughs with domainadaptive language models". In: Proceedings of the 10th ACM Conference on Web Science, pp. 177–184. ACM, May 2018

[8] S. Fana, R. Y.K.Laub and J. Leon Zhaob. 2015. "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix". Big Data Research, 2(2015), 28

[9] S. Lessmann, J. Haupt, K. Coussement, K. W. De Bock, 2019, "Targeting customers for profit: An ensemble learning framework to support marketing decision-making", https://doi.org/10.1016/j.ins.2019.05.027

[10] P. Łady ˙zy ´nski, K. ˙Zbikowski, P. Gawrysiak, 2019, "Direct marketing campaigns in retail banking with the use of deep learning and random forests", Expert Systems With Applications 134 (2019) 28–35

[11] Armin Lawi, Ali Akbar Velayaty and Zahir Zainuddin(2017). "On identifying potential direct marketing consumers using adaptive boosted support vector machine": IEEE 4th International Conference on Computer Applications and Information Processing Technology (CAIPT). https://doi.org/10.1109/CAIPT.2017.8320691

[12] McKinsey & Company. 2013. "Smart analytics can tap up to 20% of lost ROI", <www.mckinsey.com/business-functions/marketing-and-sales/our-insights/smart-analytics-can-tap-up-to-20-of-lost-roi> (2013)

[13] eMarketer. 2014. "Personalization Sees Payoffs in Marketing Emails (2014)"

[14] S. Erevelles, N. Fukawa and L. Swayne. 2016. "Big Data consumer analytics and the transformation of marketing", Journal of Business Research, 69 (Feb. 2016), 897-904

[15] A. Lemmens and C. Croux (2006). "Bagging and Boosting Classification Trees to Predict Churn": Journal of Marketing Research 43(2). http://dx.doi.org/10.1509/jmkr.43.2.276

[16] M. Andersson, M. Fredriksson and A. Berndt. 2014. "Open or Delete: Decision-makers' Attitudes Towards E-mail Marketing Messages", Advances in Social Sciences Research Journal, 1 (2014)

[17] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic and C. Huang . 2011. "Towards street-level client-independent IP geolocation", Proceedings of the 8th USENIX conference on Networked Systems Design and Implementation, (2011), 365–379

[18] B. Eriksson, P. Barford, B. Maggs and R. Nowak .2012. "Posit: a lightweight approach for IP geolocation". In: Newsletter ACM SIGMETRICS Performance Evaluation Review Archive, 40 (2012), 2–11

[19] P. Chintagunta, D. M. Hanssens, J. R. Hauser, "Marketing Science and Big Data", MARKETING SCIENCE Vol. 35, No. 3, (May–June 2016), pp. 1–2 ISSN 0732-2399 (print) I ISSN 1526-548X (online)

[20] A. A. S. Gunawan, Tania, D. Suhartono 2016. "Developing recommender systems for personalized email with big data", Conference Paper • October 2016 DOI: 10.1109/IWBIS.2016.7872893

[21] HUSAM Y, AHMAD SAMED AL, AMRO AL. "Digital marketing adoption among smes in jordan: a mixed method approach". Journal of Theoretical and Applied Electronic Commerce Research. 28th February 2019. Vol.97. No 4.

[22] DR. TAREQ N. HASHEM. "The reality of internet of things (IOT) in creating a data-driven marketing opportunity: mediating role of customer relationship management (CRM)". Journal of Theoretical and Applied Information Technology. 2019 May 1;14(2):15-32. 2021 January 31: Vol.99. No 2